DOCUMENT RESUME

ED 461 922 EA 031 515

AUTHOR Ananda, Sri; Rabinowitz, Stanley

TITLE Building a Workable Accountability System: Key Decision Points for Policymakers & Educators. Knowledge Brief.

INSTITUTION WestEd, San Francisco, CA.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 2001-00-00
NOTE 14p.
CONTRACT ED-01-CO-0012
PUB TYPE Information Analyses (070) -- Reports - Descriptive (141)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; *Educational Assessment; Elementary Secondary Education; *Evaluation Criteria; Organizational Objectives; Outcomes of Education; *Program Design; Program Effectiveness; Program Evaluation; Program Validation

ABSTRACT

Educational reform is becoming synonymous with accountability. However, designing an effective accountability system poses serious challenges for states as is clear from the progress and pitfalls experienced by states such as Texas, Kentucky, and California. For example, many states tend to overload their new accountability programs, resulting in overly complex systems with questionable reliability, validity, and fairness. This Knowledge Brief is aimed at policymakers and educators now in the process of designing or redesigning accountability programs. It identifies seven key questions that must be addressed in planning an accountability system and lays out the issues, options, and potential pitfalls relevant to each. These questions are: (1) What are the primary goals to be accomplished with an accountability system? (2) What indicators should be included in the system? (3) Which students should be included in the system and when should they be tested? (4) What is the most appropriate accountability model for a given situation? (5) What consequences can the accountability model support? (6) How can the intended and unintended effects of the accountability system be evaluated? and (7) What will be done about the problems uncovered through the accountability system? Models currently used in Texas, Kentucky, and California are briefly described. (RT)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 461 922

Building a Workable Accountability System: Key Decision Points for Policymakers & Educators

Knowledge Brief

Sri Ananda and Stanley Rabinowitz

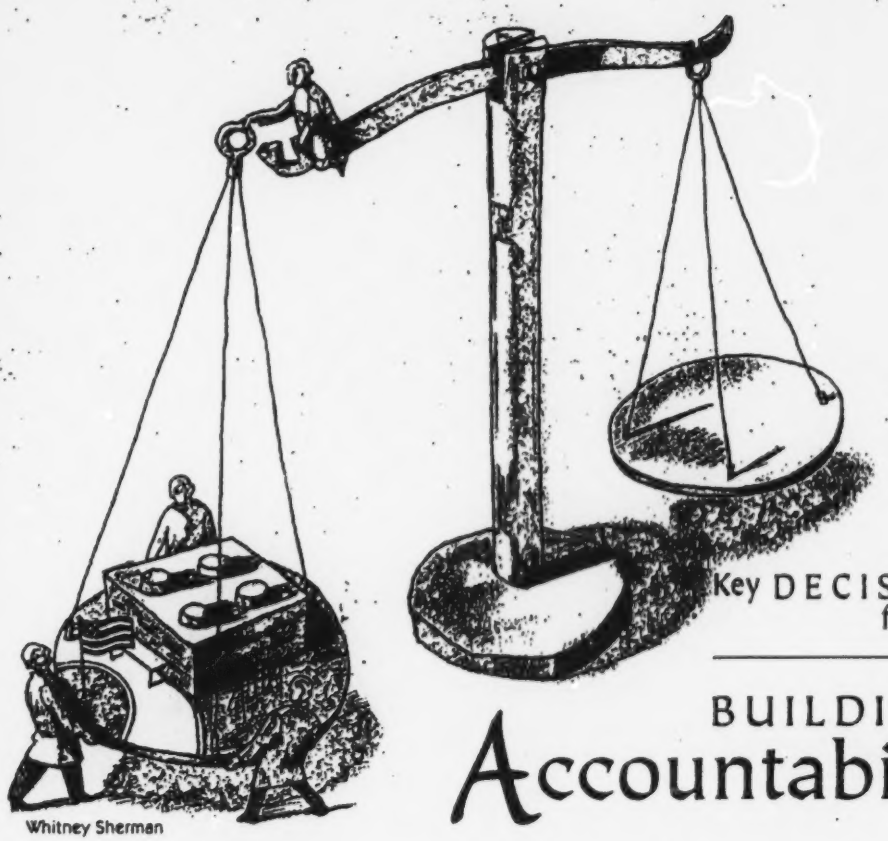
U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

EA 031515

knowledge

EA



Whitney Sherman

Key DECISION POINTS for Policymakers & Educators

BUILDING A WORKABLE Accountability System

Education reform is becoming synonymous with accountability. To a greater extent than ever before, states are relying on accountability measures to ensure that their reform efforts take hold. Furthermore, the ways in which these states are monitoring the performance of students and schools differ significantly from the past. The primary measures by which schools are being held accountable have shifted from inputs (e.g., ratio of certified staff to students, per-pupil expenditures) to outcomes (i.e., student achievement). At the same time, performance expectations are increasing and results yield official consequences, inciting teachers and administrators to do all they can to demonstrate improvements in student achievement.

The aims are laudable, yet designing an effective accountability system poses serious challenges for states. This is clear from the progress and pitfalls experienced by a number of "reforming" states, such as Kentucky, Texas, and California. For example, many states tend to overload their new accountability programs, resulting in overly complex systems with questionable reliability, validity, and fairness. Moreover, advocates of such

Written by
Sri Ananda and
Stanley Rabinowitz

WestEd
Improving education through research, development, and service

programs often expect immediate evidence of student and school performance growth. In turn, many schools that feel pressured by the state and the general public to meet achievement goals are tempted to adopt short-term "quick fixes," rather than adopting strategies aimed at long-term systemic reform. In some instances, for example, teachers focus primarily on teaching test-taking skills rather than on improving instruction.

> Ideally, accountability systems should rely on more than one indicator for evaluating school performance. But this rule of thumb is easy to violate.

States that are just starting to develop or update their accountability systems can profit from the experiences of those states where accountability efforts are well underway. This Knowledge Brief is aimed at policymakers and educators now in the process of designing or redesigning accountability programs. At the risk of oversimplifying, the brief identifies a sequence of key questions that must be addressed in planning an accountability system, then lays out the issues, options, and potential pitfalls relevant to each.

1 What are the primary goals you are trying to accomplish with an accountability system?

A common mistake is weighing down an accountability system with too many goals and targeted areas for improvement. Common sense dictates that the more goals and targeted areas of improvement that are identified, the less likely you are to achieve complete success in any of them. In fact, too broad a scope may actually inhibit attainable results as schools are forced to give attention to too many areas at once. Education reform efforts have demonstrated again and again that positive change tends to take hold not across the board, but in specific "pockets," such as in certain content areas or with certain student populations. Thus, change efforts should be focused on a reasonable number of targets.

That said, many states feel pressured to achieve multiple goals with their accountability systems. In response, they often embrace multiple sweeping education improvement goals, such as improving student learning, motivating teachers and students, reducing achievement disparity between majority and minority students, monitoring education costs, improving access to education, building public confidence in education, and improving the state's competitive economic status as compared to other states. On top of such broadly stated education goals, some states also identify goals in a more functional manner (e.g., raise test scores, show early progress in specific content areas, move indicators) for their accountability systems (Baker, 2000).

In the face of pressure to have a comprehensive accountability system, those involved in the planning must find ways to reach consensus on a few key purposes and targeted areas for improvement. Achieving consensus will inevitably be an iterative process. Typically, the process starts as a "blue sky" exercise, with those involved at the beginning of the planning process (e.g., legislature, State Board of Education, educators, community and business representatives) identifying an assortment of possible short-term and long-term goals and areas of need. As more information becomes available (e.g., expectations about when demonstrable results will be expected; what

financial and human resources can be garnered to support the system), planners should shorten and prioritize the list accordingly. One way to achieve this is to consider a phased-in accountability system with, for example, the first phase focusing on reading and writing literacy and a second phase focusing on mathematics achievement. Once performance goals for these content areas are met, additional content areas (e.g., science, workplace readiness) can be added to the accountability system. In all instances, a system's goals and targeted improvement areas should be revisited at different stages of system design and implementation to ensure that they are being adequately addressed and are still appropriate given new developments, such as shifts in leadership or education priorities.

> An effective accountability system is efficient, providing comprehensive measurement yet leaving sufficient time for academic learning to occur.

2 What indicators should be included in the system?

A fair and effective accountability system includes multiple indicators, of which there are two categories. The first is assessment — measures of student achievement and gains. The second category is non-assessment — elements perceived to influence student achievement, such as attendance and retention, or outcome measures other than test data, such as percentages of graduates enrolled in postsecondary education or employed in the workforce.

With respect to assessment indicators, the first decision is whether to include norm-referenced tests (NRTs), criterion-referenced tests (CRTs), or both. Many state accountability systems include both because they want to gauge how their students are progressing relative to state standards (for which CRTs are needed) and the status of their students relative to those in the nation as a whole (for which NRTs are needed). More complex yet is the question of whether non-traditional assessments (e.g., portfolios and other non-multiple-choice measures) are desirable and affordable.

According to a recent study by the Education Commission of the States (1999), the most commonly used indicators in a statewide accountability system are:

- Assessment scores (41 states)
- Dropout rate (33 states)
- Student attendance (29 states)
- Expenditures and use of resources (27 states)
- Graduation rate (18 states)
- Student behavior — discipline, truancy, suspension, expulsion, etc. (18 states)
- Transition to higher education or employment (16 states)

Although there is general agreement that accountability systems should rely on more than one indicator for evaluating school performance, this rule of thumb is easy to violate due to unforeseen circumstances or unrealistic planning. For example, California identified several assessment and non-assessment indicators for its Accountability Performance Index. However, at this early point of implementation, many of the indicators (high school exit examination, retention rates, dropout rates) are not yet fully developed or

are insufficiently reliable to include in the system. Thus, while the state's accountability plan calls for multiple indicators, it currently uses just one indicator: student scores on the SAT-9 examination. Judging school performance and issuing high-stakes rewards and sanctions according to a single indicator leaves a state's accountability system vulnerable to charges of unfairness and inadequacy.

3 Which students should be included in the system and when should they be tested?

Testing all grades versus selected grades. One approach to a comprehensive accountability system is to test every student in each academic content area at every grade level. Indeed, many states require the administration of norm-referenced tests in all grades for all students. This may be excessive because the evidence is that for the majority of students, NRT results tend to remain stable across adjacent grade levels. More importantly, over-testing of students is a problem. Consider the burden on a typical high school junior. In many states, these students might expect to take a high school graduation test; a nationally developed NRT; multiple state-developed CRTs linked explicitly to state content standards; state-developed end-of-course examinations in several academic and career subjects; one or more college entrance examinations (e.g., SAT, ACT); and teacher-developed classroom tests.

> A real danger with systematically excluding the scores of certain student groups from school accountability analyses is that schools may pay less attention to these students.

By contrast, an effective accountability system is efficient, providing comprehensive measurement yet leaving sufficient time for academic learning to occur. To this end, states may choose a more targeted approach, electing to assess different content areas at different grade levels (e.g., test English/language arts in grades 3, 6, and 9 and mathematics in grades 4, 7, and 10). Similarly, states could elect to focus on criterion-referenced testing at some grades and norm-referenced testing at other grades. In fact, for a number of reasons, one could argue that NRTs are more valuable at the elementary grade levels than at the secondary levels. For one thing, NRTs represent a cost-efficient way to help ensure that students do not fall too far behind, relative to basic skills development. Also, because content standards across states tend to be more similar at the elementary level than in later grades, states are more likely to find an NRT for the lower grades that is sufficiently aligned to their standards. Finally, other assessment tools are usually already available at the secondary level to give a good picture of student performance, such as high school exit exams and end-of-course exams.

Including the scores of new students. Decisions about whom and at what grades to assess must be based on explicit and fair policies. Some states specify that students be enrolled for a specific amount of time before their scores are included for school accountability purposes. For example, Wisconsin's and California's policy for assessing school performance in a given year is to exclude the test scores of students who were enrolled in that school for less than one full academic year prior to testing (Education Commission of the States, 2000). Such policies are primarily intended to protect a given school, ensuring that it is not held accountable for the performance of students who have been in its

"system" for less than a year. The possible downside of such a policy, however, is that students who tend to move around a lot are left out of the accountability indices and, therefore, might be ignored. Policymakers need to find a balance that is fair to all student populations and to the schools that serve them.

> The accountability model in Texas holds all schools to a common standard. In contrast, Kentucky's model looks at **changes** in performance.

Testing and inclusion for special populations. Federal law (e.g., Title I of the Elementary and Secondary Education Act) requires testing of students with disabilities and of English Language Learners (ELLs). Thus, accountability systems that do not have fair and inclusive policies regarding testing of these students put the state at risk of losing federal funding. Although many states now have policies about the inclusion of students with special needs, data are sorely lacking about the actual participation and inclusion rates of these students. Most states leave the decision about how best to assess special education students to students' Individualized Education Program (IEP) committees. Such committees also determine whether students require accommodations in order to participate (e.g., presentation format, such as Braille, large print, reading aloud; test administration setting; timing/scheduling). But state policies determine if the accommodation selected for a given student renders his or her score comparable to the scores of other students. If it is deemed comparable, the student's score is then counted in the accountability program.[1]

Responsibility for decisions about whether to include or exclude ELLs from accountability-driven assessments is shifting. Such decisions were once addressed primarily at the local level. Increasingly, they are addressed at the state level, and there is much variation across states concerning allowable accommodations. The decision about inclusion is most often based on how long the student has been in the United States, the amount of time he or she has been in an English-as-a-Second-Language (ESL) or bilingual program, and/or how he or she scores on a test of English proficiency. Common accommodations made for ELL students include presentation format (reading aloud, interpretation, translation of test directions and/or test items); test administration setting; and timing/scheduling.

A real danger with systematically excluding the scores of certain student groups from school accountability analyses (e.g., students who are new to the school, special education students, and ELL students) is that schools may pay less attention to these students. The result may be a less than adequate education for many students. Of course, sanctions for individual students based on their scores should be carefully considered. Equally important, to help narrow the achievement gap, accountability systems should base school rewards and sanctions on the performance growth of *all* groups of students.

4 What accountability model best serves your purposes?

No matter what high-level statistical methods a state may use in its accountability system, determining which accountability design or model to use is a fundamental decision. Linn (2000) provides a good review of some of the options available. The specifics of individual options notwithstanding, there are two basic

approaches that can be used for school accountability:

- comparing a school's current student performance data to absolute performance standards established by the state (the most commonly used approach); and
- examining a school's overall performance growth or "gains" over time (through cross-sectional or longitudinal analyses).

Below is a brief description of the models currently used by three states: Texas, Kentucky, and California.

TEXAS

Texas's accountability model focuses on the current status of a school, with all schools held to a common standard. The base indicators for accountability ratings include: Texas Assessment of Academic Skills (TAAS) in reading, writing, and mathematics; dropout rate; and attendance rate. In 1999, district and school performance was reported in four categories:

Exemplary — at least 90% of all students and each student group — African American, Hispanic, White, and economically disadvantaged — must pass each section of the TAAS;

Recognized — at least 80% of all students and each student group must pass each section of the TAAS;

Academically Acceptable/Acceptable — at least 45% of all students and each student group must pass each section of the TAAS; and

Academically Unacceptable/Low-Performing — not meeting the standards for Academically Acceptable or higher and not achieving required improvement in identified low-performing areas.

> States committed to closing the achievement gap should seriously consider disaggregating scores.

As shown in its rating categories, the Texas system calls for schools to keep track of performance by student groups and assigns schools to performance categories based both on the performance of all students and on the performance of these targeted groups. States committed to closing the achievement gap should seriously consider disaggregating scores and designating school performance levels in this way.

Even though the Texas model holds all schools accountable to the same performance standard, since 1995 the state has raised the standard for acceptable performance each year. One strength of this approach is its recognition that, practically speaking, achievement of high performance standards is a long-term process. Raising the bar over time allows schools the opportunity to systematically implement curriculum and instructional changes needed to support higher student achievement. In the short term, however, a state may find itself in the position of having to defend its policy of rewarding schools that meet only the interim bar rather than a standard the public would more likely find acceptable. For example, the public may consider a 45 percent pass rate unacceptable, preferring to say that nothing less than a 60 percent rate should be considered acceptable. But a state that initially sets high pass rates may find that virtually none of its schools meet the standard, as happened in the first year of Virginia's accountability program, eroding public confidence in the system.

The Texas model may not be appropriate in cases where schools vary significantly in terms of student

performance levels because it may unfairly penalize schools that demonstrate reasonable progress, but do not yet meet the common performance standards.

KENTUCKY

In contrast to Texas, Kentucky's accountability model looks at *changes* in performance, based on comparisons of student cohorts across grades. Also, rather than directly assigning schools to categories, Kentucky uses an index approach, or formula, to assign schools a numerical value that shows how well they are performing along a continuum. Kentucky's "accountability index" combines a school's academic factors (i.e., student performance on assessments in several traditional and non-traditional content areas) and nonacademic factors (e.g., increasing attendance rates, decreasing retention and dropout rates, improving the transition to adult life).

> There is disagreement as to the efficacy of rewards and sanctions in public education. They may produce changes in practice, but the question is whether such changes are permanent or transitory.

Kentucky's model uses a two-year accountability cycle, with schools required to meet growth goals based on their baseline performance. For example, the 1996-97 and 1997-98 school years served as the baseline for each school, against which progress was then assessed for the 1998-99 and 1999-2000 school years. Combining data across years in this way is a good strategy. It improves the reliability of an accountability system, and it promotes greater public confidence in any decisions based on the results. However, even using multiple years of data may not result in sufficiently reliable information to make fine distinctions between categories of school performance. Schools may vary in their placement due as much to variations in student populations across years as to actual classroom practice.

In the Kentucky system, each school is assigned to one of five performance categories, based on the school's accountability index:

Meets Goal — meets or exceeds its predicted performance for the accountability cycle;

Maintaining (Dropout Not Met) — while the school's accountability index meets or exceeds expectations, the dropout rate is not sufficiently low to meet established standards;

Maintaining — the accountability index is less than its predicted performance and greater than the Assistance point for the accountability cycle;

Assistance — score in the top two thirds of the schools classified as Assistance, based on the school's final accountability index; and

Assistance Audit — scored in the bottom one third of schools classified as Assistance based on their final accountability index.

Although Kentucky's model sets growth goals for individual schools based on their baseline, it has also set a common goal for all schools by the end of 20 years. Thus, over the long run, schools that have started out with a low accountability index are expected to demonstrate faster growth rates than those schools that initially score well. This approach represents an interesting compromise between using an absolute standard against which to judge each school and allowing schools to demonstrate growth from their individual baseline performances.

CALIFORNIA

Like Kentucky, California uses a performance growth approach. Based on student performance on the annual SAT-9 exam, California ranks schools into categories (1-10, with 10 as the highest) for each grade span of instruction (elementary, middle, and high school). A second 1-to-10 ranking indicates how each school compares against others with similar socioeconomic characteristics. The plan is that by June 2001, school rankings will reflect the annual growth rate targeted for each school, its actual growth rate, and how its growth rate compares to schools with similar characteristics (pupil mobility; ethnicity and socioeconomic status; percentage of fully credentialed teachers; percentage of ELLs; average class size per grade level; whether schools operate year-round programs).

California's school accountability system has several noteworthy features. Like Texas, California reports assessment results by student groups (special education students; ELLs; minority groups), signaling California's commitment to closing the achievement gap. Including two sets of rankings for each school is another interesting feature; it highlights a school's relative standing compared against all other schools, as well as its relative standing in comparison to others with similar student characteristics. As previously mentioned, however, California's current use of only one indicator — the SAT-9 exam scores — to rank schools (because the other planned indicators are not yet ready) is a significant flaw at this stage of implementation.

5 What consequences can your accountability system support?

Rewards and sanctions are key components in the accountability systems for a number of states. Specifically, rewards are made to teachers and schools based on attainment of performance goals, while sanctions may be applied against individual schools (or students) when student achievement or progress falls below set standards.

It appears that states are paying insufficient attention to ensuring validity and reliability in their accountability systems.

There is disagreement as to the efficacy of rewards and sanctions in public education. Although such consequences may produce changes in practice (e.g., more targeted instructional support to low-performing students), the question remains as to whether such changes are permanent or transitory (Education Commission of the States, 1999). Moreover, there is the issue of fairness: Are the rewards and sanctions based on valid and reliable indicators?[2] Unreliable systems may lead to inconsistent and inaccurate classifications of schools. For example, a given school may be classified as eligible for rewards one year and be identified for sanctions the next, due to fluctuations in the value of the school's accountability indices. Such fluctuations raise questions about fairness of an accountability system because it defies common sense that the performance of a given school can vary so significantly from one accountability cycle to the next. This situation actually occurred in Kentucky, where some schools were eligible for rewards in one cycle and sanctions in the next. The situation caused confusion in the field and questions about the credibility of the overall accountability system.

At this point, it appears that states are paying insufficient attention to ensuring validity and

reliability in their accountability systems. To avoid the situation faced by Kentucky and other states, a state should carefully investigate both the validity and reliability of its accountability system, ensuring that differences in school performance categories (or indices) reflect accurate and meaningful differences in accomplishment or growth. Generally speaking, the more indicators and data points that are incorporated into an accountability system, the greater its reliability (Hill, 2000).

6 How can the intended and unintended effects of the accountability system be evaluated?

Intended consequences of an accountability system might include improved instructional practices, increased student performance and learning, and increased public support for schools. Unintended consequences might include widespread cheating on high-stakes tests, increased student dropout rates, and negative public outcry if, for example, large numbers of schools fail to meet improvement targets. To better ensure intended consequences and minimize unintended consequences, a state must carefully and systematically monitor its accountability system — starting even before the system becomes operational and continuing for its duration.

Unintended consequences of an accountability system might include widespread cheating on high-stakes tests and increased student dropout rates.

In ongoing evaluation of the accountability system, the following questions should be asked:

- Are the long-term and short-term goals of the system worthwhile, realistic, and achievable?
- To what degree does the system support high-quality instruction and student access to education; minimize corruption; affect teacher quality; and produce unanticipated outcomes?
- What are the actual costs incurred by the system and what are the necessary trade-offs between quality and cost?
- What support (e.g., professional development) do teachers and administrators need to implement the system?
- How will parents and the general public be informed as to the goals and limitations of the system?

The need for ongoing evaluation is underscored by the negative attention some state accountability systems are now receiving. The popular media are filled with stories about states with ambitious school accountability plans that have to make just-in-time policy retreats because of public outcry (Baker, 2000). Such unintended consequences might have been avoided or minimized through more careful planning and ongoing evaluation of the accountability system.

For example, in Massachusetts, this year's 10th graders were to be the first class required to pass the new Massachusetts Comprehensive Assessment System (MCAS) examinations in order to graduate. Yet the state's Commissioner of Education recently recommended that students who fail the exams be allowed to: (1) earn local

certificates anyway; and (2) take scaled-down versions of the exam, which would be designed solely to determine whether students have met minimum passing standards on the original exam (Gehring, 2000). Such policy retreats can cut into the credibility of an accountability effort, undermining public support. In the case of Massachusetts, a *Boston Globe* editorial labeled the Commissioner of Education's proposed re-test idea as "MCAS Lite," calling it a retreat from high standards. This development is just the latest in the heated debate in Massachusetts over the MCAS. Public support for the state's accountability system had already begun to wane months earlier when hundreds of high school students from a dozen or so schools boycotted the test because they felt it was unduly difficult and, therefore, unfair. Similar public debates about the fairness of high-stakes testing are underway in Arizona, Nevada, California, Virginia, and many other states where the deadlines for sanctions are approaching.

Besides policy retreats, another unintended consequence can be an increase in dropout rates. FairTest (2000), a Cambridge, Massachusetts-based advocacy group (and vehement opponent of standardized testing), recently issued a report claiming that more students are dropping out of school in Massachusetts, in part because of the testing program. Similarly, Texas is experiencing an increase in dropout rates among certain minority groups, an increase that some critics attribute to the TAAS high school graduation examination. The critics point to the differential TAAS high school graduation "pass rates" between Whites and minorities as a contributing factor to the differential dropout rates.

Still another unanticipated consequence for a number of state accountability systems is corruption. It is most frequently exhibited by cheating and inappropriate teaching to the test.

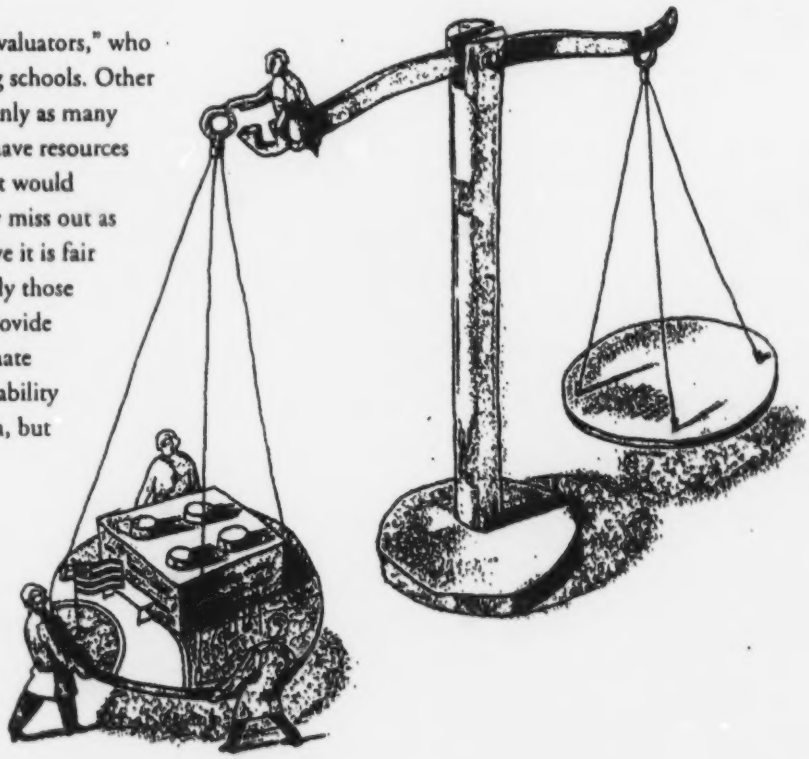
7 What will you do about the problems uncovered through the accountability system?

Certainly, every state should strive for a well-designed accountability system that is sufficiently valid and reliable to support sanctions and rewards and is adequately monitored for intended and unintended consequences. However, even that is not enough. States have an obligation to help fix the problems highlighted by their accountability system, including providing technical assistance and financial support to low-performing schools. For example, Virginia Governor James Gilmore recently announced that his state will provide $1.2 million to 189 schools with very low scores on the state assessments in English and mathematics. Schools must use the money to provide more instruction time in English and mathematics, using any instructional approach they deem appropriate. In addition, the state is sending "academic review teams" to all schools that received state warnings regarding their accountability performance. The teams, composed of retired teachers and education specialists, will work with teachers, principals, and superintendents to develop a plan for bringing each school up to full accreditation standards. This is similar to Kentucky's use of "Distinguished Educators," who are assigned to low-performing schools, as well as

States have an obligation to help fix the problems highlighted by their accountability system, including providing technical assistance and financial support to low-performing schools.

to California's use of "External Evaluators," who assist California's low-performing schools. Other states, such as Nevada, identify only as many low-performing schools as they have resources to assist. While some schools that would benefit from extra assistance may miss out as a result, state policymakers believe it is fair to identify as low-performing only those schools to whom the state can provide extra support. After all, the ultimate goal of a comprehensive accountability system is not to reward or punish, but to improve student learning.

The authors would like to thank David Berliner, Richard Hill, Erica Adelsheimer, and Paul Koehler for their valuable feedback on this Knowledge Brief.



For a complete listing of WestEd's publications, visit our Web site at WestEd.org

ENDNOTES

1 While states have expended increasing energy and resources in determining how to include special education students into state assessment programs, much less thought has gone into how best to include the results of these assessments into accountability systems. Traditionally, state accountability initiatives have excluded the results of special education students. However, this approach can lead to questionable practices (e.g., inappropriate classifications of students as handicapped to avoid accountability) and is inconsistent with federal regulations. Two general approaches are available to states as they consider how to include their special education students into accountability systems. First, they may choose to hold such students to the same standards as their non-classified counterparts. However, this approach has two shortcomings: (1) it may force students to master standards not included in their IEP; and (2) it requires accommodated or alternate assessments to provide results equivalent to the mainstream state assessment, a difficult technical feat to accomplish. On the other hand, states may consider special education students "successful" if they meet the specific standards of their IEP, *even if these differ from the state content or performance standards*. Such an approach may be considered fairer to individual special education students (and the schools they attend) but may require states to endorse differential levels of achievement for different populations of students.

2 Researchers are beginning to distinguish between reliable and valid *assessment data* versus reliable and valid *accountability systems*. Although the foundation of a valid accountability system is a valid assessment program, a valid assessment program does not assure a valid accountability system. For an in-depth discussion of technical issues related to state assessment and accountability systems, see Baker (2000) and Hill (2000).

REFERENCES

Baker, E. *Improving the validity of accountability systems* (presentation as part of the Edward F. Reidy, Jr., Lecture Series: Technical Accountability Systems, at The National Center for the Improvement of Educational Assessment, Dover, NH: October 5-6, 2000).

Education Commission of the States. (2000). *Clearinghouse notes: accountability incentives/sanctions.* Denver, CO: www.ecs.org

Education Commission of the States. (1999). *Education accountability systems in the 50 states.* Denver, CO: www.ecs.org

FairTest (The National Center for Fair and Open Testing), CARE (Coalition for Authentic Reform in Education). (September, 2000). MCAS alert. *MCAS: Making the Massachusetts dropout crisis worse.* Cambridge, MA: www.fairtest.org

Gehring, J. (October 18, 2000). Massachusetts eyes alternate diploma as graduation test looms. *Education Week:* www.edweek.org

Hill, R. *Reliability of accountability systems* (presentation as part of the Edward F. Reidy, Jr., Lecture Series: Technical Accountability Systems, at The National Center for the Improvement of Educational Assessment, Dover, NH: October 5-6, 2000).

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher, 29 (2),* 4.

WestEd is a nonprofit research, development, and service agency that works with education and other communities to promote excellence, achieve equity, and improve learning for children, youth, and adults. Drawing on the best knowledge from research and practice, our agency collaborates with practitioners, policymakers, and others, addressing critical education and other human service issues. WestEd, with initiatives throughout the United States and abroad, is one of the nation's designated Regional Educational Laboratories — originally created by Congress in 1966 — serving the states of Arizona, California, Nevada, and Utah. With headquarters in San Francisco, WestEd has offices across the United States.

For more information about WestEd, visit our Web site: WestEd.org/; call 415.565.3000 or, toll-free, (1.877) 4-WestEd; or write:

WestEd
730 Harrison Street
San Francisco, CA 94107-1242

This report was produced in whole or in part with funds from the Office of Educational Research and Improvement, U.S. Department of Education, under contract #ED-01-CO-0012. Its contents do not necessarily reflect the views or policies of the Department of Education.

© 2001 WestEd. All rights reserved. Permission to reproduce, with WestEd copyright notice included, is hereby granted.

WestEd®
730 Harrison Street
San Francisco
California 94107-1242

Address service requested

Non-Profit
U.S. Postage
PAID
Los Alamitos, CA
90720
Permit No. 87

BEST COPY AVAILABLE